Analyzing Control-Treatment Writing Sample Comparisons: Fresno, CA iACCESS Drama Integration Spring 2017 Report

Lawrence Scripp, Principal Investigator
Josh Gilbert, Research Associate
Center for Music and the Arts in Education
May 26th, 2017

Introduction

This report will examine treatment-control differences in social studies writing topics among 5th and 6th grade students involved in the iACCESS project. Our sample included 166 pieces of writing from control and treatment students in 5th and 6th grade classrooms across various schools in Fresno, California.

Student writing samples were scored:

- First in terms of word counts for each writing sample
- Second according to a 4-point rubric that included four independent writing element measures labeled as "Conventions," "Organization/Focus," "Development/Elaboration," and "Empathy/Social-Emotional Language." These four elements were averaged to determine the mean writing score for each student.

The table below presents descriptive statistics for the 5th and 6th grade treatment and control groups.

Table 1

Grade	Research Cohort	N	Mean Writing Score	SD	Min	Max
5th	Control	51	1.69	.41	1	2.63
5th	Treatment	67	1.81	.36	1	3
6th	Control	28	1.96	.42	1.13	2.75
6th	Treatment	20	2.25	.35	1.5	2.63

The following sections of the report will examine treatment-control differences within each grade level in terms of the word count, mean writing score, and then a closer look at the four individual scoring categories of the rubric.

In addition to presenting mean differences and pooled standard deviations, we also include Cohen's *d*, a measure of "effect size," which is defined as the mean difference between groups divided by the pooled standard deviation. This measure enables us to interpret the substantive size of treatment-control differences in addition to their statistical significance (as indicated by the t-ratios and associated p-values). For the purposes of this report, we will interpret the d

values from the perspective that 0.4 constitutes a small effect, 0.7 a medium effect, and 1.0 a large effect.

Control-Treatment Differences

A. Word Count

The table below summarizes control-treatment differences by grade level for word count. In both 5th and 6th grade, the treatment students wrote substantially more than the control students, as indicated by the large and medium effect sizes of 1.01 and .61, respectively. t-test results indicate that these differences are statistically significant for both grades at the p<.05 level.

Table 2

Grade Level	Control Mean	Treatment Mean	Difference	Pooled SD	d (effect size)	t-test
5	45.8 n=52	83.0 n=67	37.3	37.4	1.00 (large)	t=5.40 p<.0001
6	69.9 n=28	89.8 n=20	19.9	32.7	.61 (medium)	t=2.10 p=.04

Mean Score (Average of conventions, organization, development, and social-emotional language)

The table below summarizes control-treatment differences by grade level for mean score ratings (the average of conventions, organization, development, and social emotional language). In both 5th and 6th grade, treatment students scored significantly higher than control students, though the magnitude of the difference was substantially larger for 6th grade (d=.74) than for 5th grade (d=.36).

Table 3

Grade Level	Control Mean	Treatment Mean	Difference	Pooled SD	d (effect size)	t-test
5	1.69 n=51	1.82 n=66	.14	.37	.36 (small)	t=1.96 p=.05
6	1.96 n=28	2.25 n=20	.29	.39	.74 (medium)	t=2.51 p=.02

The subsequent sections of the report examine treatment-control comparisons for each of the four elements that comprise the mean score: conventions, organization/focus, development/elaboration, and empathy/social emotional language.

Writing Element 1: Conventions

The table below summarizes control-treatment differences by grade level for ratings of ELA Conventions (spelling, grammar, and punctuation). In 5th grade, treatment and control average convention scores were indistinguishable. In 6th grade, however, treatment students' Conventions scores were significantly higher than those of the control school students (t=1.98, p=.05, d=.58).

Table 4

Grade Level	Control Mean	Treatment Mean	Difference	Pooled SD	d (effect size)	t-test
5	1.96 n=51	1.98 n=66	.02	.86	.01	t=.15 p=.88 (N.S.)
6	2.54 n=28	3.05 n=20	.51	.88	.58 (medium)	t=1.98 p=.05

Writing Element 2: Organization/Focus

The table below summarizes control-treatment differences by grade level for ratings of writing Organization/Focus. In both 5th and 6th grade, treatment students substantially outperformed control students in ratings for organization/focus as indicated by the moderate effect sizes of .60 and .52, respectively. Note that in 6th grade, despite the relatively robust effect size of .52, this finding is positively trending ($p<.1$) though not statistically significant ($p<.05$) due to the small sample size.

Table 5

Grade Level	Control Mean	Treatment Mean	Difference	Pooled SD	d (effect size)	t-test
5	1.87 n=50	2.17 n=66	.30	.51	.60 (medium)	t=3.18 p=.002
6	2.27 n=28	2.58 n=20	.31	.60	.52 (small)	t=1.76 p=.09

Writing Element 3: Development/Elaboration

The table below summarizes control-treatment differences by grade level for ratings of Development/Elaboration of descriptive language. In both 5th and 6th grade, treatment students substantially outperformed control students in ratings for development/elaboration as indicated by the medium and large effect sizes of .71 and .93, respectively.

Table 6

Grade Level	Control Mean	Treatment Mean	Difference	Pooled SD	d (effect size)	t-test
5	1.72 n=50	2.00 n=66	.28	.39	.71 (medium)	t=3.80 p=.0002
6	2.00 n=28	2.35 n=20	.35	.38	.93 (large)	t=3.17 p=.003

Writing Element 4: Empathy/Social-Emotional Language

The table below summarizes control-treatment differences by grade level ratings for degree of included Empathy/Social-Emotional language. The very low average scores indicate that social-emotional language was very rare throughout the writing samples, and neither 5th nor 6th grade students showed statistically significant treatment-control differences. Note that control students' scores in this category were slightly higher than those of treatment students, hence the negative (though statistically insignificant) d values and t-ratios.

Table 7

Grade Level	Control Mean	Treatment Mean	Difference	Pooled SD	d (effect size)	t-test
5	1.17 n=50	1.13 n=66	-.04	.30	-.14 (very small)	t=-.74 p=.46 (N.S.)
6	1.02 n=28	1.00 n=20	-.02	.07	-.24 (very small)	t=-.84 p=.40 (N.S.)

Overall Correlations

The matrix below presents (a) the pair-wise correlation between word count[1] and mean score, and then (b) a 4x4 correlation matrix that displays the degrees of association among word count and the four individual writing element for the entire sample of treatment and control students. Because the four writing elements of conventions, organization/focus, development/elaboration, and empathy/social emotional language were averaged to determine the "mean score," the correlations between each of these sub-categories and the overall mean score is not included in the correlation matrix.

In general, three findings stand out:

- Empathy/Social-emotional language was clearly not significantly correlated with any other variable, likely attributable the lack of use of this writing element across the data sample, which should be of concern to the goals of the iACCESS project.
- Moderate and weak associations existed between conventions and the other variables, which could be interpreted as indicating the relative independence of ELA conventions to the other expressive and organizational elements of writing content.
- Very strong associations existed among word count, development/elaboration, and organization/focus, indicating the importance of quantity (word count) needed for high levels of organization and elaboration to emerge.

Table 8: Correlation Matrix (n=164)

	Log(Word Count)	Conventions	Organization-Focus	Development-Elaboration
Mean Score	.61****			
Conventions	.27***			
Organization-Focus	.67****	.44****		
Development-Elaboration	.80****	.40****	.82****	
Social-Emotional Language	.15~	.03	.07	.09

~p<.1, *p<.05, **p<.01, ***p<.001, ****p<.0001

[1] Because word count was heavily right skewed, we applied a logarithmic transformation to normalize the distribution.

More specifically, the results of the table above show:

- The strong positive correlation between mean score and word count ($r=.61$, $p<.0001$), indicating that students who wrote more scored higher, on average.
- The very strong correlation between word count and development/elaboration scores ($r=.80$, $p<.0001$), indicating that word count is a strong predictor of student development and elaboration, explaining 64% ($r^2=.8^2=.64$) of the variance in development/elaboration.
- Conventions scores were moderately positively correlated with both organization/focus ($r=.44$, $p<.0001$) and development/elaboration ($r=.40$, $p<.0001$), suggesting that students whose writing contained fewer spelling, grammatical, and punctuation errors were likely to score higher on measures of organization and development, on average.
- Organization/focus and development/elaboration were very strongly correlated ($r=.82$, $p<.0001$), indicating that students whose writing was best organized were likely to be the most detailed, on average.
- Given that ratings for social-emotional language were so low for both treatment and control students, it is not surprising that the correlations between social emotional language and the other variables are very weak and not statistically significant.

Conclusions

Examining treatment-control writing score differences by grade level revealed that on average, treatment students scored higher than control students in virtually all levels of writing elements. Though in terms of overall mean score, treatment and control 5th graders only slightly different, subsequent analyses of the individual components of the mean score revealed that 5th grade treatment students substantially outperformed control students in the areas of Organization/Focus and Development/Elaboration.

However, this analysis is limited by the absence of additional control variables that would enable us to more precisely estimate the magnitude of treatment-control differences, such as prior academic achievement (baseline scores), and other demographics such as gender, ethnicity, SES, and so forth. Therefore, we cannot yet make conclusive statements about the potential impact of the iACCESS program.